

Multi-Terminal Source Coding With Action ¹ Dependent Side Information

Yeow-Khiang Chia, Himanshu Asnani and Tsachy Weissman

Department of Electrical Engineering, Stanford University

Email: ykchia@stanford.edu, asnani@stanford.edu, tsachy@stanford.edu

Abstract

We consider multi-terminal source coding with a single encoder and multiple decoders where either the encoder or the decoders can take cost constrained actions which affect the quality of the side information present at the decoders. For the scenario where decoders take actions, we characterize the rate-cost trade-off region for lossless source coding, and give an achievability scheme for lossy source coding for two decoders which is optimum for a variety of special cases of interest. For the case where the encoder takes actions, we characterize the rate-cost trade-off for a class of lossless source coding scenarios with multiple decoders. Finally, we also consider extensions to other multi-terminal source coding settings with actions, and characterize the rate-distortion-cost tradeoff for a case of successive refinement with actions.

I. INTRODUCTION

The problem of source coding with decoder side information (S.I.) was introduced in [1]. S.I. acts as an important resource in rate distortion problems, where it can significantly reduce the compression rate required. In classical Shannon theory and in work building on [1], S.I. is assumed to be either always present or absent. However, in practical systems as we know, acquisition of S.I. is costly, the encoder or decoder has to expend resources to acquire side information. With this motivation, the framework for the problem of source coding with action-dependent side information (S.I.) was introduced in [2], where the authors considered the cases where the encoder or decoder are allowed to take actions (with cost constraints) that affect the quality or availability of the side information present at the decoders, and in some settings, the encoder. As noted in [2], one motivation for this setup is the case where the side information is obtained via a sensor through a sequence of noisy measurements of the source sequence. The sensor may have limited resources, such as acquisition time or power, in obtaining the side information. This is therefore modeled by the cost constraint on the action sequence to be taken at the decoder. Additional motivation for considering this framework is given in [2]. We also refer readers to recent work in [3], [4] for related Shannon theoretic scenarios invoking the action framework.

In this paper, we extend the source coding with action framework to the case where there are multiple decoders, which can take actions that affect the quality or availability of S.I. at each decoder, or where the encoder takes actions that affect the quality or availability of S.I. at the decoders. As a motivation for this framework, consider the following problem: An encoder observes an i.i.d source sequence X^n which it wishes to describe to two decoders via a common rate limited link of rate R . The decoders, in addition to observing the output of the common rate limited link, also have access to a common sensor which gives side information Y that is correlated with X . However, because of contention or resource constraints, when decoder 1 observes the side information, decoder 2 cannot access the side information and vice versa. This problem is depicted in Figure 1. Even in the absence of cost constraints on the cost of switching to 1 or 2, this problem is interesting and non-trivial. How should the decoders share the side information and what is the optimum sequence of actions to be conveyed and then taken by the decoder?

By posing the above problem in the framework of source coding with action dependent side information, we solve it for the (near) lossless source coding case, a special case of lossy source coding with switching dependent side information, and give interpretations of the standard random binning and coding arguments when specialized to this switching problem. As one example for the implications of our findings, when $Y = X$, we show that the optimum rate required for lossless source coding in the above problem is $H(X)/2$ - clearly a lower bound on the required rate, but that it suffices for perfect reconstruction of the source simultaneously at both decoders is, at first glance, surprising. We devote a significant portion of this paper to the setting where the side information

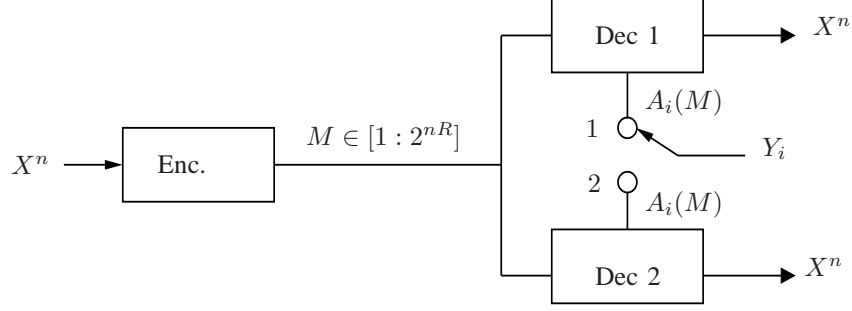


Fig. 1: Lossless source coding with switching dependent side information. When the switch is at position 1, decoder 1 observes the side information. When the switch is at position 2, decoder 2 observes the side information.

at the decoders is obtained through a switch that determines which of the two decoders gets to observe the side information, and obtain a complete characterization of the fundamental performance limits in various scenarios involving such switching. The achieving schemes in these scenarios are interesting in their own right, and also provide insight into more general cases.

The rest of the paper is organized as follows. In section II, we provide formal definitions and problem formulations for the cases considered. In section III, we first consider the setting of lossless source coding with decoders taking actions with cost constraints and give the optimum rate-cost trade-off region for this setting. Next, we consider the setting of lossy source coding decoders taking actions with cost constraints and give a general achievability scheme for this setup. We then specialize our achievability scheme to obtain the optimum rate-distortion and cost trade-off region for a number of special cases. In section V, we consider the setting where actions are taken by the encoder. The rate-cost-distortion tradeoff setting is open even for the single decoder case. Hence, we only consider a special case of lossless source coding for which we can characterize the rate-cost tradeoff. In section VI, we extend our setup to two other multiple users settings, including the case of successive refinement with actions. The paper is concluded in section VII.

II. PROBLEM DEFINITION

In this section, we give formal definitions for, and focus on, the case where there are two decoders. Generalization of the definitions to K decoders is straightforward, and, as we indicate in subsequent sections, some of our results hold in the K decoders setting. We follow the notation of [5]. We use A to denote the action random variable. The distortion measure between sequences is defined in the usual way. Let $d : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow [0, \infty)$. Then, $d(x^n, \hat{x}^n) := \frac{1}{n} \sum_{i=1}^n d(x_i, \hat{x}_i)$. The cost constraint is also defined in the usual fashion: let $\Lambda(A^n) := \frac{1}{n} \sum_{i=1}^n \Lambda(A_i)$. Throughout this paper, sources (X^n, Y^n) are specified by the joint distribution $p(x^n, y^n) = \prod_{i=1}^n p_{X,Y}(x_i, y_i)$ (i.i.d.). The decoders obtain side information through a discrete memoryless *action channel* $P_{Y_1, Y_2 | X, A}$ specified by conditional distribution $p(y_1^n, y_2^n | x^n, a^n) = \prod_{i=1}^n p_{Y_1, Y_2 | X, A}(y_{1i}, y_{2i} | x_i, a_i)$, with decoder j obtaining side information Y_j^n for $j \in \{1, 2\}$. Extensions to more than two sources or more than two channel outputs for multiple decoders are straightforward.

A. Source coding with actions taken at the decoders

This setting for two decoders is shown in figure 2. A $(n, 2^{nR})$ code for the above setting consists of one encoder

$$f : \mathcal{X}^n \rightarrow M \in [1 : 2^{nR}],$$

one joint action encoder at all decoders

$$f_{A-Dec.} : M \in [1 : 2^{nR}] \rightarrow \mathcal{A}^n,$$

and two decoders

$$\begin{aligned} g_1 : \mathcal{Y}_1^n \times [1 : 2^{nR}] &\rightarrow \hat{\mathcal{X}}_1^n, \\ g_2 : \mathcal{Y}_2^n \times [1 : 2^{nR}] &\rightarrow \hat{\mathcal{X}}_2^n, \end{aligned}$$

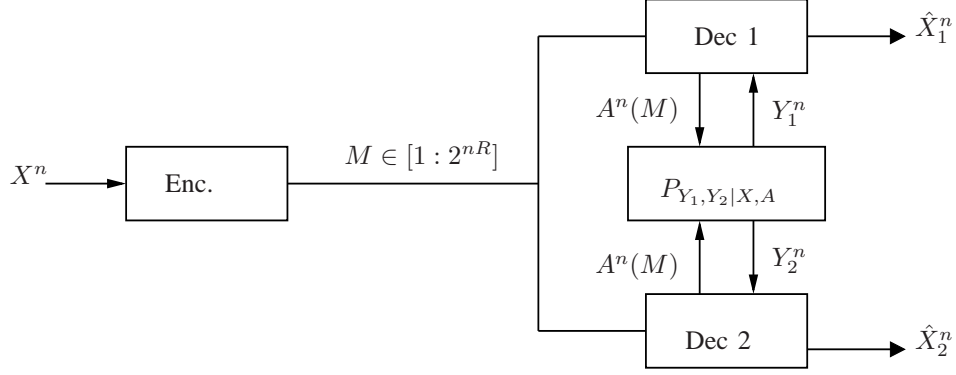


Fig. 2: Lossy source coding with actions at the decoders.

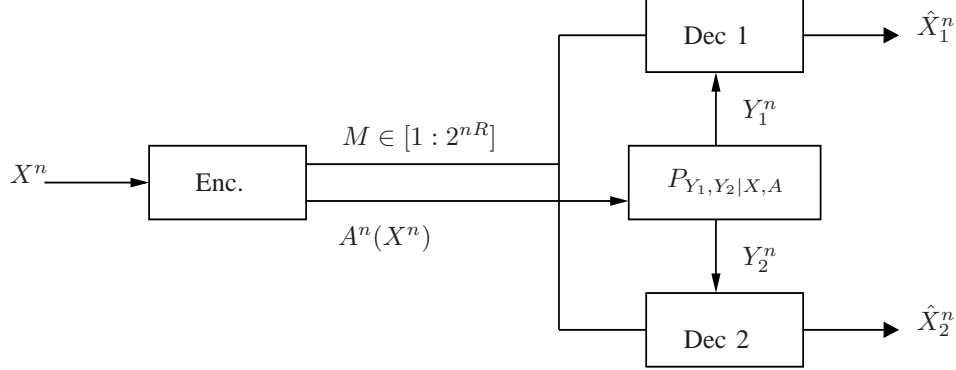


Fig. 3: Lossy source coding with actions at the encoder.

Given a distortion-cost tuple (D_1, D_2, C) , a rate R is said to be achievable if, for any $\epsilon > 0$ and n sufficiently large, there exists $(n, 2^{nR})$ code such that

$$\begin{aligned} \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n d_j(X_i, \hat{X}_{j,i}) \right] &\leq D_j + \epsilon, \quad j=1,2, \\ \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \Lambda(A_i) \right] &\leq C + \epsilon. \end{aligned}$$

The *rate-distortion-cost region*, $\mathcal{R}(D_1, D_2, C)$, is defined as the infimum of all achievable rates.

Causal reconstruction with action dependent side information: Some results in this paper involves the case of *causal reconstruction*. In the case of causal reconstruction, the decoder reconstructs \hat{X}_i based only on the received message M and the side information up to time i . That is,

$$g_{j,i} : \mathcal{Y}_j^i \times [1 : 2^{nR}] \rightarrow \hat{\mathcal{X}}_{j,i},$$

for $j \in \{1, 2\}$ and $i \in [1 : n]$.

Remark 2.1: The case of the decoders taking separate actions A_1 and A_2 respectively is a special case of our setup since we can write $A := (A_1, A_2)$.

Remark 2.2: For the reconstruction mappings, we excluded the action sequence as an input since A^n is a function of the other input M . In our (information) rate expressions, we will see the appearance of A in the expressions. As we will see in the next subsection, an advantage of this definition is that it carries over to the case when the encoder takes actions rather than the decoders.

B. Source coding with action taken at the encoder

This setting is shown in figure 3. As the definitions and problem statement for this case are similar to the first setting, we will only mention the differences between the two settings. The main difference is that the encoder takes actions rather than the decoders. Therefore, in the definition of a code, we replace the case of a joint action encoder at the decoders with the *encoder taking actions* given by the function

$$f_{A-Enc} : \mathcal{X}^n \rightarrow \mathcal{A}^n.$$

As in the setting of actions taken at the decoder, here too we assume that the side information observed by the decoders is not available at the encoder. In subsequent sections we also describe the results pertaining to the case where side information is available at the encoder.

Remark 2.3: Lossless source coding - Some of our results concern the case of lossless source coding. In the case of lossless source coding, the definitions are similar, except that the distortion constraints D_1, D_2 are replaced by the block probability of error constraint: $P(\{\hat{X}_1^n \neq X^n\} \cup \{\hat{X}_2^n \neq X^n\}) \leq \epsilon$.

III. LOSSLESS SOURCE CODING WITH ACTIONS AT THE DECODERS

In this section and the next, we consider the case of source coding with actions taken at the decoders. We first present results for the lossless source coding setting. While the lossless case can be taken to be a special case of lossy source coding, we present them separately, as we are able to obtain stronger results for more general scenarios in the lossless setting, and give several interesting examples that arise from this setup. The case of lossy source coding for two decoders is presented in section IV.

For the lossless case, we first state the result for the general case of K decoders. Our result is stated in Theorem 1.

Theorem 1: Let the action channel be given by the conditional distribution $P_{Y_1, Y_2, \dots, Y_K | X, A}$ with decoder j observing the side information Y_j . Then, the minimum rate required for lossless source coding with actions taken at the decoders and cost constraint C is given by

$$R = \min \max_{j \in [1:K]} \{H(X|Y_j, A)\} + I(X; A),$$

where min is taken over the distributions $p(x)p(a|x)p(y_1, y_2, \dots, y_K|x, a)$ such that $E \Lambda(A) \leq C$.

Achievability

As the achievability techniques used are fairly standard (cf. [5]), we give only a sketch of achievability.

Codebook Generation:

- Generate $2^{n(I(X;A)+\epsilon)}$ A^n sequences according to $\prod_{i=1}^n p(a_i)$.
- Bin the set of all X^n sequences into $2^{n(\max_{j \in [1:K]} \{H(X|Y_j, A)\} + \epsilon)}$ bins, $\mathcal{B}(m_b)$, $m_b \in [1 : 2^{n(\max_{j \in [1:K]} \{H(X|Y_j, A)\} + \epsilon)}]$.

Encoding:

- Given a source sequence x^n , the encoder looks for an index $M_A \in [1 : 2^{n(I(X;A)+\epsilon)}]$ such that $(x^n, a^n(M_A)) \in \mathcal{T}_\epsilon^{(n)}$. If there is none, it outputs a uniform random index from $[1 : 2^{n(I(X;A)+\epsilon)}]$. If there is more than one such index, it selects an index uniformly at random from the set of feasible indices. From the covering lemma [5, Chapter 3], the probability of error for this step goes to 0 as $n \rightarrow \infty$ since there are $2^{n(I(X;A)+\epsilon)}$ A^n sequences.
- The encoder also looks the index $m_b \in [1 : 2^{n(\max_{j \in [1:K]} \{H(X|Y_j, A)\} + \epsilon)}]$ such that $x^n \in \mathcal{B}(m_b)$.
- It then sends the indices m_b and M_A to the decoders via the common link. This step requires a rate of $R = \max_{j \in [1:K]} \{H(X|Y_j, A)\} + I(X; A) + 2\epsilon$.

Decoding:

- The decoders take the joint action $a^n(M_A)$ and obtain their side informations Y_j for $j \in [1 : K]$.
- Decoder j then looks for the *unique* X^n sequence in bin $\mathcal{B}(m_b)$ such that $(X^n, Y_j^n, a^n(M_A)) \in \mathcal{T}_\epsilon^{(n)}$. An error is declared if there is none more than one x^n sequence satisfying the decoding condition. The probability of error for this step goes to 0 as $n \rightarrow \infty$ from the strong law of large numbers and the fact that $|\mathcal{B}| > 2^{n(\max_{j \in [1:K]} \{H(X|Y_j, A)\})}$.

Converse

Given a $(n, 2^{nR}, C)$ code, consider the rate constraint for decoder j . We have

$$\begin{aligned}
nR &\geq H(M) \\
&= I(M; X^n) \\
&\stackrel{(a)}{=} I(A^n; X^n) + I(M; X^n | A^n) \\
&\stackrel{(b)}{\geq} I(A^n; X^n) + H(M | A^n, Y_j^n) - H(M | A^n, X^n, Y_j^n) \\
&= H(X^n) - H(X^n | A^n) + I(M; X^n | A^n, Y_j^n) \\
&\stackrel{(c)}{\geq} H(X^n) - H(X^n | A^n) + H(X^n | A^n, Y_j^n) - n\epsilon_n \\
&= H(X^n) - H(X^n | A^n) + H(X^n | A^n) \\
&\quad + H(Y_j^n | X^n, A^n) - H(Y_j^n | A^n) - n\epsilon_n \\
&\stackrel{(d)}{\geq} \sum_{i=1}^n H(X_i) + H(Y_{ji} | X_i, A_i) - H(Y_{ji} | A_i) - n\epsilon_n.
\end{aligned}$$

(a) follows from A^n being a function of M ; (b) follows from the Markov chain $M \rightarrow (X^n, A^n) \rightarrow Y_j^n$; (c) follows from the assumption of lossless source coding; (d) follows from conditioning reduces entropy and the fact that the action channel is a discrete memoryless channel (DMC). Define Q as the standard time sharing random variable. Observe that $H(X_Q | Q) = H(X_Q) = H(X)$, $H(Y_{jQ} | A_Q, X_Q, Q) = H(Y_{jQ} | A_Q, X_Q) = H(Y_j | A, X)$ and $H(Y_{jQ} | A_Q, Q) \leq H(Y_j | A)$. Hence, we can write the lower bound as

$$\begin{aligned}
nR &\geq n(H(X) + H(Y_j | X, A) - H(Y_j | A) - \epsilon_n) \\
&= n(I(X; A) + H(X | Y_j, A) - \epsilon_n).
\end{aligned}$$

Taking the intersection of all lower bounds for all K decoders then give us the rate expression given in the Theorem. Finally, the cost constraint on the action follows from $C \geq \mathbb{E} \frac{1}{n} \sum_{i=1}^n \Lambda(A_i) = \mathbb{E} \Lambda(A)$.

We now specialize the result in Theorem 1 to the case of source coding with switching dependent side information mentioned in the introduction. We consider the more general setting involving K decoders.

Corollary 1: Source coding with switching dependent side information and no cost constraints. Let (X, Y) be jointly distributed according to $p(x, y)$. Let $\mathcal{A} = [1 : K]$ and $P_{Y_1, Y_2, \dots, Y_K | X, A}$ be defined by $Y_j = Y$ when $A = j$ and e otherwise for $j \in [1 : K]$. Let $\Lambda(A) := 0$ for all $a \in \mathcal{A}$. Then, the minimum rate is given by

$$H(X | Y) + \frac{K-1}{K} I(X; Y).$$

Proof:

Proof of Corollary 1 amounts to an explicit characterization of the distribution of $p(a|x)$ in Theorem 1. For each $j \in [1 : K]$, we have, from Theorem 1,

$$\begin{aligned}
R &\geq H(X | Y_j, A) + I(X; A) \\
&= H(X | Y) + I(X; Y) - I(X; Y_j | A).
\end{aligned} \tag{1}$$

Consider now the sum

$$\begin{aligned}
\sum_{j=1}^K I(X; Y_j | A) &\stackrel{(a)}{=} \sum_{a \in \mathcal{A}} p(a) I(X; Y | A = a) \\
&= H(Y | A) - H(Y | X, A) \\
&\stackrel{(b)}{\leq} H(Y) - H(Y | X) \\
&= I(X; Y).
\end{aligned} \tag{2}$$

(a) follows from the fact that $Y_j = e$ for $a \neq j$ and $Y_j = Y$ for $a = j$. (b) follows from the Markov Chain $A - X - Y$.

Next, summing over the K lower bounds in (1), we obtain

$$\begin{aligned} R &\geq \frac{1}{K}(KH(X|Y) + KI(X;Y) - \sum_{j=1}^K I(X;Y_j|A)) \\ &\geq H(X|Y) + I(X;Y) - \frac{1}{K}I(X;Y) \\ &= H(X|Y) + \frac{K-1}{K}I(X;Y), \end{aligned}$$

where we used inequality (2) in the second last step. Finally, noting that this lower bound on the achievable rate can be obtained from Theorem 1 by setting $A \perp X$ and $p(a = j) = 1/K$ completes the proof of Corollary 1. \blacksquare

Remark 3.1: The action can be set to a fixed sequence independent of the source sequence. This is perhaps not surprising since there is no cost on the actions.

Remark 3.2: For $K = 2$ and $X = Y$, which is the example given in the introduction, we have $R = H(X)/2$.

Remark 3.3: For this class of channels, the achievability scheme in Theorem 1 has a simple and interesting “modulo-sum” interpretation. We present a sketch of an alternative scheme for this class of switching channels for $K = 2$. It is straightforward to extend the achievability scheme given below to K decoders.

Alternative achievability scheme

Split the X^n sequence into 2 equal parts; $X_1^{n/2}$ and $X_{n/2+1}^n$ and select the fixed action sequence of letting decoder 1 observe $Y_1^{n/2}$ and decoder 2 observe $Y_{n/2+1}^n$. Separately compress each part using standard random binning with side information to obtain $M_1 \in [1 : 2^{n(H(X|Y)/2+\epsilon)}]$ and $M_2 \in [1 : 2^{n(H(X|Y)/2+\epsilon)}]$ corresponding to the first and second half respectively. Within each bin, with high probability, there are only $2^{nI(X;Y)/2}$ typical $X^{n/2}$ sequences and we represent each of them with an index $M_{j1} \in [1 : 2^{n(I(X;Y)/2+\epsilon)}]$, where $j \in \{1, 2\}$. Send out the indexes M_1 and M_2 , which requires a rate of $H(X|Y) + 2\epsilon$. Next, send out the index $M_{11} \oplus M_{21}$ which requires a rate of $I(X;Y)/2 + \epsilon$. From M_1 and side information $Y_1^{n/2}$, decoder 1 can recover $X_1^{n/2}$ with high probability. Therefore, it can recover M_{11} with high probability. Hence, it can recover M_{21} from $M_{11} \oplus M_{21}$ and therefore, recover the $X_{n/2+1}^n$ sequence. The same analysis holds for decoder 2 with the indices interchanged.

Corollary 2 gives the characterization of the achievable rate for a general switching dependent side information setup with cost constraint on the actions for two decoders.

Corollary 2: General switching dependent side information for 2 decoders. Define the action channel as follows: $A \in \{0, 1, 2, 3\}$; $A = 0, Y_1 = e, Y_2 = e$; $A = 1, Y_1 = Y, Y_2 = e$; $A = 2, Y_1 = e, Y_2 = Y$; and $A = 3, Y_1 = Y, Y_2 = Y$. Let $\Lambda(A = j) = C_j$ for $j \in [0 : 3]$. Then, the optimum rate-cost trade-off for this class of channel is given by

$$\begin{aligned} R &\geq I(X; A) + \max\{H(X|Y_1, A), H(X|Y_2, A)\} \\ &= I(X; A) + p_0 H(X|A = 0) + \sum_{j=1}^3 p_j H(X|Y, A = j) \\ &\quad + \max\{p_1 I(X; Y|A = 1), p_2 I(X; Y|A = 2)\}, \end{aligned}$$

for some $p(a|x)$, where $P\{A = j\} = p_j$, satisfying $\sum_{j=0}^3 p_j C_j \leq C$.

Remark 3.4: This setup again has a “modulo-sum interpretation” for the term $\max\{p_1 I(X; Y|A = 1), p_2 I(X; Y|A = 2)\}$ and the rate can also be achieved by extending the achievability scheme described in Corollary 1. The scheme involves partitioning the X^n sequence according to the value of A_i for $i \in [1 : n]$. Following the scheme in Corollary 1, we let $M_j \in [1 : 2^{n(p_j H(X|Y, A=j)+\epsilon)}]$ for $j \in [0 : 3]$. We first generate a set of A^n codewords according to $\prod_{i=1}^n p(a_i)$. Next, for each A^n codeword, define A^{n_j} to be $\{A_i : A_i = j\}$. Similarly, let $X^{n_j} := \{X_i : A_i = j, i \in [1 : n]\}$ be the set of possible X sequences corresponding to A^{n_j} . We bin the set of all X^{n_j} sequences to $2^{n(p_j H(X|Y, A=j)+\epsilon)}$ bins, $\mathcal{B}_j(M_j)$. For $j \in \{1, 2\}$, further bin the set of x^{n_j} sequences into $2^{n(p_j I(X; Y|A=j)+\epsilon)}$ bins, $\mathcal{B}_{j1}(M_{j1})$, $M_{j1} \in [1 : 2^{n(p_j I(X; Y|A=j)+\epsilon)}]$.

For encoding, given an x^n sequence, the encoder first finds an A^n sequence that is jointly typical with x^n . It sends out the index corresponding to the A^n sequence found. Next, it splits the x^n sequence into four partial

sequences, x^{n_j} , for $j \in [0 : 3]$, where x^{n_j} is the set of x_i corresponding to $A_i = j$. It then finds the corresponding bin indices such that $x^{n_j} \in \mathcal{B}_j(I_j)$ for $j \in [0 : 3]$. It then sends out the indices M_0, M_1, M_2, M_3 and $M_{11} \oplus M_{21}$.

For decoding, we mention only the scheme employed by the first decoder, since the scheme is the same in for decoder 2. From the properties of jointly typical sequences and standard analysis for Slepian-Wolf lossless source coding [6], it is not difficult to see that decoder 1 can recover $x^{n_0}, x^{n_1}, x^{n_3}$ with high probability. Recovery of x^{n_1} also allows decoder 1 to recover the index M_{11} and hence, M_{21} from $M_{11} \oplus M_{21}$. Noting that the rate of M_{21} and M_2 sums up to $p_2 H(X|A=2) + 2\epsilon$, it is then easy to see that decoder 1 can recover x^{n_2} with high probability.

In corollary 1, we showed that, for the case of switching dependent side information, the action sequence is independent of the source X^n when cost constraint on the actions is absent. A natural question to ask is whether the action is still independent of X^n when a cost constraint on the actions is present? The following example shows that the optimum action sequence is in general dependent on X^n .

Example 1: Action is dependent on source statistics when cost constraint is present. Let $K = 2$ and (X, Y) be distributed according to an S channel, with $X \sim \text{Bern}(1/2)$, $P(Y = 1|X = 1) = 1$ and $P(Y = 0|X = 0) = 0.2$. Let $A \in \{1, 2\}$ with $Y_1 = Y$ if $A = 1$ and $Y_2 = Y$ if $A = 2$. Let $P(A = 1) = p_1$, $P(X = 0|A = 1) = 1/2 + \delta_1$ and $P(X = 0|A = 2) = 1/2 - \delta_2$. Figure 4 shows the probability distributions between the random variables.

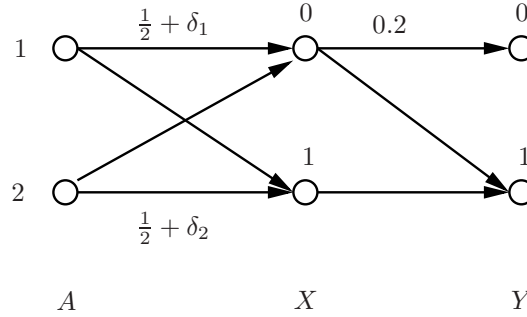


Fig. 4: Probability distributions for random variables used in example 1

Since $X \sim \text{Bern}(1/2)$, δ_1 and δ_2 are related by $\delta_2 = p_1 \delta_1 / (1 - p_1)$. Therefore, we set $\delta_1 = \delta$ and $\kappa = p_1 / (1 - p_1)$ for this example.

Now, let $\Lambda(A = 1) = 1$ and $\Lambda(A = 2) = 0$ and $C = 0.4$. The optimum rate-cost tradeoff in this case may be obtained from Corollary 2 by setting $C_0 = C_3 = \infty$, $C_1 = 1$ and $C_2 = 0$, giving us

$$R = I(X; A) + p_1 H(X|Y, A = 1) + (1 - p_1) H(X|Y, A = 2) \\ + \max\{p_1 I(X; Y|A = 1), (1 - p_1) I(X; Y|A = 2)\},$$

for some $p(a|x)$, where $P\{A = 1\} = p_1$, satisfying $p_1 \leq 0.4$. The problem of finding the optimum action sequence to take then reduces (after some straightforward algebra) to the following optimization problem:

$$\min_{p_1, \delta} 1 - p_1 H_2(0.5 - \delta) - (1 - p_1) H_2(0.5 - \kappa \delta) \\ + p_1 H(X|Y, A = 1) + (1 - p_1) H(X|Y, A = 2) \\ + \max\{p_1 I(X; Y|A = 1), (1 - p_1) I(X; Y|A = 2)\}, \\ \text{subject to} \\ 0 \leq p_1 \leq 0.4, \\ -0.5 \leq \delta \leq 0.5,$$

where

$$\begin{aligned}
H(X|A=1) &= p_1 H_2(0.5 - \delta), \\
H(X|A=2) &= (1 - p_1) H_2(0.5 - \kappa\delta), \\
H(X|Y, A=1) &= ((0.5 + \delta)(0.8) + 0.5 - \delta) H_2 \left(\frac{0.5 - \delta}{(0.5 + \delta)(0.8) + (0.5 - \delta)} \right), \\
H(X|Y, A=2) &= ((0.5 - \kappa\delta)(0.8) + (0.5 + \kappa\delta)) H_2 \left(\frac{0.5 + \kappa\delta}{(0.5 - \kappa\delta)(0.8) + (0.5 + \kappa\delta)} \right),
\end{aligned}$$

and $H_2(\cdot)$ is the binary entropy function.

While exact solution to this (non-convex) optimization problem involves searching over p_1 and δ , it is easy to see that if A is restricted to be independent of X , which corresponds to restricting δ to be equal to 0, then the optimum solution for p_1 is 0.4. Under $p_1 = 0.4$ and $\delta = 0$, we obtain $R_{A \perp X} = 0.9568$. In contrast, setting $p_1 = 0.4$ and $\delta = -0.05$, we obtain $R = 0.9554$, which shows that the optimum action sequence is in general dependent on the source X when cost constraints are present.

An explanation for this observation is as follows. The cost constraint forces decoder 1 to see less of the side information Y than decoder 2. It may therefore make sense to bias the distribution $X|A=1$ so that Y conveys more information about the source sequence X , even at the expense of describing the action sequence to the decoders. Roughly speaking, the amount of information conveyed about X by Y may be measured by $I(X; Y)$. Note that under $\delta = 0$, $I(X; Y|A=1) = 0.108$, whereas under $\delta = -0.05$, $I(X; Y|A=1) = 0.1116$. A plot of the optimum rate versus cost tradeoff obtained by searching over a grid of p_1 and δ is shown in Figure 5. The figure also shows the rate obtained if actions were forced to be independent of the source sequence.

IV. LOSSY SOURCE CODING WITH ACTION AT THE DECODERS

In this section, we first consider the case when causal reconstruction is required, and give the general rate-distortion-cost region for K decoders. Next, we consider the case of lossy noncausal reconstruction for two decoders and give a general achievability scheme for this case. We then show that our achievability scheme is optimum for several special cases. Finally, we discuss some connections between our setting and the complementary delivery setting introduced in [7].

A. Causal reconstruction for K decoders

Theorem 2: Causal lossy reconstruction for K decoders

When the decoders are restricted to causal reconstruction [8], $\mathcal{R}(D_1, D_2, \dots, D_K, C)$ is given by

$$R = I(U; X)$$

for some $p(u|x)$, $A = f(U)$ and reconstruction functions \hat{x}_j for $j \in [1 : K]$ such that

$$\begin{aligned}
\mathbb{E} d_j(X, \hat{x}_j(U, Y_j)) &\leq D_j \text{ for } j \in [1 : K] \\
\mathbb{E} \Lambda(A) &\leq C.
\end{aligned}$$

The cardinality of U is upper bounded by $|U| \leq |\mathcal{X}||\mathcal{A}| + K$.

Remark 4.1: Theorem 2 generalizes the corresponding result for one decoder in [2, Theorem 3].

Proof: As the achievability scheme is a straightforward extension of the scheme in [2, Theorem 3], we will omit the proof of achievability here. For the converse, given a code that satisfies the cost and distortion constraints,

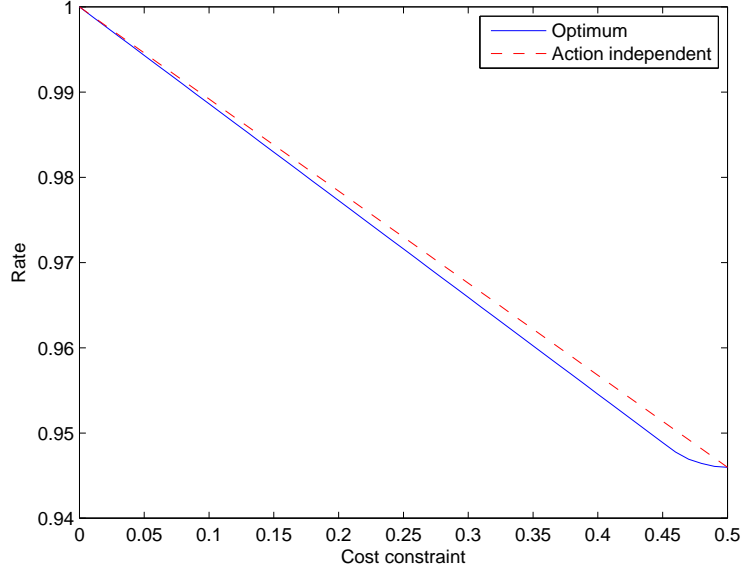


Fig. 5: Rate versus cost constraint for the example 1. It is easy to show operationally that the optimum rate versus cost curve is convex in the cost constraint. When the cost constraint approaches zero, the rate approaches 1, since this case corresponds to decoder 1 not seeing any of the side information. When the cost constraint approaches 0.5, the rate approaches the minimum rate without cost constraint. The red dashed line shows the rate that would be obtained if actions were forced to be independent of the source. As can be seen on graph, forcing actions to be independent of the source is in general not optimum when cost constraint is present. The optimum rate versus cost constraint plot appears to be linear over a range of cost constraints. It can be shown that if the cost constraint is below a threshold, then the optimum rate is a linear function of the cost constraint. However, the plot obtained via numerical simulation appears to be linear in the cost constraint over a wider range than what we obtained by analysis. Performing a more refined analysis to obtain a cost constraint threshold that matches the cost threshold obtained by simulation appears to be difficult, due to the nature of the optimization problem that is involved.

we have

$$\begin{aligned}
nR &\geq H(M) \\
&= I(X^n; M) \\
&\stackrel{(a)}{=} \sum_{i=1}^n (H(X_i) - H(X_i|M, X^{i-1})) \\
&\stackrel{(b)}{=} \sum_{i=1}^n (H(X_i) - H(X_i|M, X^{i-1}, A^{i-1})) \\
&\stackrel{(c)}{=} \sum_{i=1}^n (H(X_i) - H(X_i|M, X^{i-1}, A^{i-1}, Y_1^{i-1}, \dots, Y_K^{i-1})) \\
&\geq \sum_{i=1}^n (H(X_i) - H(X_i|U_i)),
\end{aligned}$$

where (a) follows from the fact that X^n is a memoryless source; (b) follows from the fact that A^{i-1} is a function of M ; (c) follows from the fact that the action channel $p(y_1, y_2, \dots, y_k|x, a)$ is a memoryless channel; and the last step follows from defining $U_i = (M, Y_1^{i-1}, \dots, Y_K^{i-1})$. Finally, defining Q to be a random variable uniform over

$[1 : n]$, independent of all other random variables, $U = (U_Q, Q)$, $X = X_Q$, $A = A_Q$ and $Y_j = Y_{jQ}$ for $j \in [1 : K]$ then gives the required lower bound on the minimum rate required. Further, we have $A = f(U)$. It remains to verify that the cost and distortion constraints are satisfied. Verification of the cost constraint is straightforward. For the distortion constraint, we have for $j \in [1 : K]$

$$\mathbb{E} \frac{1}{n} \sum_{i=1}^n d_j(X_i, \hat{x}_{ji}(M, Y_j^i)) \geq \mathbb{E} d_j(X, \hat{x}'_j(U, Y_j)),$$

where we define $\hat{x}'_j(U, Y_j) := \hat{x}_{jQ}(M, Y_j^i)$. This shows that the definition of the auxiliary random variable U satisfies the distortion constraints. Finally, the cardinality of U can be upper bounded by using the support lemma [9]. We require $|\mathcal{X}||\mathcal{A}| - 1$ letters to preserve $P_{X,A}$, which also preserves the cost constraint. In addition, we require $K + 1$ letters to preserve the rate and K distortion constraints. ■

We now turn to the case of noncausal reconstruction. For this setting, we give results only for the case of two decoders.

B. Noncausal reconstruction for two decoders

We first give a general achievability scheme for this setting.

Theorem 3: An achievable scheme for the lossy source coding with actions at the decoders is given by

$$\begin{aligned} R \geq & I(X; A) + \max \{I(X; U|A, Y_1), I(X; U|A, Y_2)\} \\ & + I(X; V_1|U, A, Y_1) + I(X; V_2|U, A, Y_2) \end{aligned}$$

for some $p(x)p(a|x)p(u|a, x)p(v_1|u, a, x)p(v_2|u, a, x)p(y_1, y_2|x, a)$ and reconstruction functions \hat{x}_1 and \hat{x}_2 satisfying

$$\begin{aligned} \mathbb{E} d_j(X, \hat{x}_j(U, V_j, A, Y_j)) &\leq D_j \text{ for } j = 1, 2, \\ \mathbb{E} \Lambda(A) &\leq C. \end{aligned}$$

We provide a sketch of achievability in Appendix A since the techniques used are fairly straightforward. As an overview, the encoder first tells the decoders the action sequence to take. It then sends a common description of X^n , U^n , to both decoders. Based on the action sequence A^n and the common description U^n , the encoder sends V_1^n and V_2^n to decoders 1 and 2 respectively. We do not require decoder 1 to decode V_2^n , or for decoder 2 to decode V_1^n .

Theorem 3 is optimum for the following special cases.

Proposition 1: Heegard-Berger-Kaspi [10], [11] Extension. Suppose the following Markov chain holds: $(X, A) - (A, Y_1) - (A, Y_2)$. Then, the rate-distortion-cost trade-off region is given by

$$\begin{aligned} R \geq & I(X; A) + I(X; U|A, Y_2) \\ & + I(X; V_1|U, A, Y_1) \end{aligned}$$

for some $p(x)p(a|x)p(u, v_1|x, a)p(y_1|x, a)p(y_2|y_1, a)$ satisfying

$$\begin{aligned} \mathbb{E} d_1(X, \hat{X}_1(U, V_1, A, Y_1)) &\leq D_1, \\ \mathbb{E} d_2(X, \hat{X}_2(U, A, Y_2)) &\leq D_2, \\ \mathbb{E} \Lambda(A) &\leq C. \end{aligned}$$

The cardinality of the auxiliary random variables is upper bounded by $|\mathcal{U}| \leq |\mathcal{X}||\mathcal{A}| + 2$ and $|V_1| \leq |\mathcal{U}|(|\mathcal{X}||\mathcal{A}| + 1)$. The achievability for this proposition follows from Theorem 3 by setting $V_2 = \emptyset$ and noting that since $(X, A) - (A, Y_1) - (A, Y_2)$, the terms in the $\max\{\cdot\}$ function simplifies to $I(X; U|A, Y_2)$. We give a proof of converse as follows.

Converse: Given a code that satisfies the constraints,

$$\begin{aligned} nR &\geq H(M) \\ &= H(M, A^n) \end{aligned}$$

$$\begin{aligned}
&= H(A^n) + H(M|A^n) \\
&\geq H(A^n) - H(A^n|X^n) + H(M|A^n, Y_2^n) - H(M|Y_2^n, A^n, X^n) \\
&= I(X^n; A^n) + I(X^n; M|A^n, Y_2^n) \\
&= I(X^n; A^n) + I(X^n; M, Y_1^n|A^n, Y_2^n) - I(X^n; Y_1^n|M, A^n, Y_2^n) \\
&= I(X^n; A^n) + H(X^n|A^n, Y_2^n) - H(X^n|M, Y_1^n, A^n, Y_2^n) - I(X^n; Y_1^n|M, A^n, Y_2^n) \\
&= I(X^n; A^n) + H(X^n|A^n, Y_2^n) - \sum_{i=1}^n (H(X_i|M, Y_1^n, A^n, Y_2^n, X^{i-1}) + I(X^n; Y_{1i}|M, A^n, Y_2^n, Y_1^{i-1})) \\
&\geq I(X^n; A^n) + H(X^n|A^n, Y_2^n) - \sum_{i=1}^n (H(X_i|M, Y_1^n, A^n, Y_2^n) + I(X^n; Y_{1i}|M, A^n, Y_2^n, Y_1^{i-1})) \\
&\stackrel{(a)}{=} I(X^n; A^n) + H(X^n|A^n, Y_2^n) - \sum_{i=1}^n (H(X_i|M, Y_1^n, A^n, Y_2^n) + I(X_i; Y_{1i}|M, A^n, Y_2^n, Y_1^{i-1})) \\
&= I(X^n; A^n) + H(X^n|A^n, Y_2^n) - \sum_{i=1}^n H(X_i|M, Y_1^{i-1}, A^n, Y_2^n) \\
&\quad + \sum_{i=1}^n (I(X_i; Y_{1i}^n|M, A^n, Y_2^n, Y_1^{i-1}) - I(X_i; Y_{1i}|M, A^n, Y_2^n, Y_1^{i-1})) \\
&= I(X^n; A^n) + H(X^n|A^n, Y_2^n) - \sum_{i=1}^n H(X_i|M, Y_1^{i-1}, A^n, Y_2^n) \\
&\quad + \sum_{i=1}^n (I(X_i; Y_{1,i+1}^n|M, A^n, Y_2^n, Y_1^i), \\
&= I(X^n; A^n) + H(X^n|A^n, Y_2^n) - \sum_{i=1}^n H(X_i|M, Y_1^{i-1}, A^n, Y_2^n) \\
&\quad + \sum_{i=1}^n (I(X_i; Y_{1,i+1}^n|M, A^n, Y_2^{n \setminus i}, Y_{1i}, Y_1^{i-1}),
\end{aligned}$$

where (a) follows from the fact that $X^{n \setminus i} - (M, A^n, Y_2^n, Y_1^{i-1}, X_i) - Y_{1i}$ and the last step follows from the Markov Chain assumption $X_i - (A_i, Y_{1i}) - (A_i, Y_{2i})$. Consider now,

$$\begin{aligned}
I(X^n; A^n) + H(X^n|A^n, Y_2^n) &= I(X^n; A^n) + H(X^n, Y_2^n|A^n) - H(Y_2^n|A^n) \\
&= H(X^n) + H(Y_2^n|A^n, X^n) - H(Y_2^n|A^n) \\
&\geq \sum_{i=1}^n (H(X_i) + H(Y_{2i}|X_i, A_i) - H(Y_{2i}|A_i)).
\end{aligned}$$

Hence,

$$\begin{aligned}
nR &\geq \sum_{i=1}^n (H(X_i) + H(Y_{2i}|X_i, A_i) - H(Y_{2i}|A_i)) - \sum_{i=1}^n H(X_i|M, Y_1^{i-1}, A^n, Y_2^n) \\
&\quad + \sum_{i=1}^n (I(X_i; Y_{1,i+1}^n|M, A^n, Y_2^n, Y_{1i}, Y_1^{i-1})).
\end{aligned}$$

Define now Q to be a random variable uniform over $[1 : n]$, independent of all other random variables; $X = X_Q$, $Y_1 = Y_{1Q}$, $Y_2 = Y_{2Q}$, $A = A_Q$, $U_i = (M, Y_1^{i-1}, A^{n \setminus i}, Y_2^{n \setminus i})$, $V_i = Y_{1,i+1}^n$, $U = (U_Q, Q)$ and $V = V_Q$. Then, we

have

$$\begin{aligned}
R &\geq H(X) + H(Y_2|X, A) - H(Y_2|A, Q) - H(X|A, Y_2, U) \\
&\quad + I(X; V|A, Y_1, U) \\
&\geq H(X) + H(Y_2|X, A) - H(Y_2|A) - H(X|A, Y_2, U) \\
&\quad + I(X; V|A, Y_1, U) \\
&= I(X; A) + I(X; U|A, Y_2) + I(X; V|A, Y_1, U).
\end{aligned}$$

It remains to verify that the definitions of U , V and A satisfy the distortion and cost constraints, which is straightforward. Prove of the cardinality bounds follows from standard techniques. ■

The next proposition extends our results for the case of switching dependent side information to the a class of lossy source coding with switching dependent side information.

Proposition 2: Special case of switching dependent side information. Let $Y_1 = X, Y_2 = Y$ if $A = 1$ and $Y_1 = Y, Y_2 = X$ if $A = 2$ and for all x , there exists \hat{x}_1 and \hat{x}_2 such that $d_1(x, \hat{x}_1) = 0$ and $d_2(x, \hat{x}_2) = 0$. Then, the rate-distortion-cost trade-off region is given by

$$\begin{aligned}
R &\geq I(X; A) + \max\{P(A = 2)I(X; U_1|A = 2, Y), \\
&\quad P(A = 1)I(X; U_2|A = 1, Y)\}
\end{aligned}$$

for some $p(x, y)p(a|x)p(u_1|x, a = 2)p(u_2|x, a = 1)$ satisfying

$$\begin{aligned}
P(A = 2) \mathbb{E} d_1(X, \hat{X}_1(Y, U_1)|A = 2) &\leq D_1, \\
P(A = 1) \mathbb{E} d_2(X, \hat{X}_2(Y, U_2)|A = 1) &\leq D_2, \\
\mathbb{E} \Lambda(A) &\leq C.
\end{aligned}$$

The cardinality of the auxiliary random variables is upper bounded by $|\mathcal{U}_1| \leq |\mathcal{X}| + 1$ and $|\mathcal{U}_2| \leq |\mathcal{X}| + 1$. Achievability follows from Theorem 3 by setting $V_1 = V_2 = \emptyset$ and $U = U_2$ if $A = 1$ and $U = U_1$ if $A = 2$. We give the proof of converse as follows.

Converse: Given a code that satisfies the cost and distortion constraints, consider the rate required for decoder 1. We have

$$\begin{aligned}
nR &\geq H(M) \\
&= H(M, A^n) \\
&= H(A^n) + H(M|A^n) \\
&\geq H(A^n) - H(A^n|X^n) + H(M|A^n, Y_1^n) - H(M|Y_1^n, A^n, X^n) \\
&= I(X^n; A^n) + I(X^n; M|A^n, Y_1^n) \\
&= I(X^n; A^n) + H(X^n, Y_1^n|A^n) - H(Y_1^n|A^n) - H(X^n|M, A^n, Y_1^n) \\
&= H(X^n) + H(Y_1^n|A^n, X^n) - H(Y_1^n|A^n) - H(X^n|M, A^n, Y_1^n) \\
&\geq \sum_{i=1}^n (H(X_i) + H(Y_{1i}|X_i, A_i) - H(Y_{1i}|A_i)) - \sum_{i=1}^n H(X_i|M, A^n, Y_1^n).
\end{aligned}$$

As before, we define Q to be an uniform random variable over $[1 : n]$, independent of all other random variables. We then have

$$\begin{aligned}
R &\geq H(X_Q|Q) + H(Y_{1Q}|X_Q, A_Q, Q) - H(Y_Q|A_Q, Q) - H(X_Q|M, A^n, Y_1^n, Q) \\
&\stackrel{(a)}{\geq} H(X) + H(Y_1|X, A) - H(Y_1|A) - H(X|M, A^n, Y_1^n) \\
&\stackrel{(b)}{=} I(X; A) + I(X; U_1|Y_1, A).
\end{aligned}$$

(a) follows from the discrete memoryless nature of the action channel and the fact that conditioning reduces entropy;

(b) follows from defining $U_{1i} = (M, A^{n \setminus i}, Y_1^{n \setminus i})$ and $U_1 = (U_{1Q}, Q)$. Expanding the second term in terms of A

and using the observation that $Y_1 = X$ when $A = 1$ and $Y_1 = Y$ when $A = 2$, we obtain

$$R \geq I(X; A) + P(A = 2)I(X; U_1|Y, A = 2).$$

For decoder 2, the same steps with side information Y_2 instead of Y_1 and defining $U_{2i} = (M, A^{n \setminus i}, Y_2^{n \setminus i})$, $U_2 = (U_{2Q}, Q)$ yield

$$R \geq I(X; A) + P(A = 1)I(X; U_2|Y, A = 1).$$

Taking the maximum over two lower bounds yield

$$R \geq I(X; A) + \max\{P(A = 2)I(X; U_1|Y, A = 2), P(A = 1)I(X; U_2|Y, A = 1)\}$$

for some $p(a|x)p(u_1, u_2|x, a)$. Verifying the cost constraint is straightforward. As for the distortion constraint, we have for the decoder 1

$$\begin{aligned} \frac{1}{n} \mathbb{E} d_1(X^n, \hat{x}_1^n(M, A^n, Y_1^n)) &= \mathbb{E} d_1(X, \hat{x}_1(U_1, A, Y_1)) \\ &= P(A = 2) \mathbb{E}(d_1(X, \hat{x}_1(U_1, Y))|A = 2). \end{aligned}$$

The same arguments hold for decoder 2. It remains to show that the probability distribution can be restricted to the form $p(a|x)p(u_1|a, x)p(u_2|a, x)$. Observe that $P(A = 2) \mathbb{E}(d_1(X, \hat{x}_1(U_1, Y))|A = 2)$ and $P(A = 2)I(X; U_1|Y, A = 2)$ depends on the joint distribution only through the marginal $p(a, u_1|x)$ and $P(A = 1) \mathbb{E}(d_2(X, \hat{x}_2(U_2, Y))|A = 1)$ and $P(A = 1)I(X; U_2|Y, A = 1)$ depends on the joint distribution only through the marginal $p(a, u_2|x)$. Hence, restricting the joint distribution to the form $p(a|x)p(u_1|a, x)p(u_2|a, x)$ does not affect the rate, cost or distortion constraints. It remains to bound the cardinality of the auxiliary random variables used, which follows from standard techniques. This completes the proof of converse. \blacksquare

Remark 4.2: The condition on the distortion constraints is simply to remove distortion offsets. It can be removed in a fairly straightforward manner.

Remark 4.3: As with the lossless source coding with switching dependent side information case, a modulo sum interpretation for the terms in the max expression is possible. When $A = 1$, the encoder codes for decoder 2, resulting, after binning, in an index I_2 for the codeword U_2^n ; and when $A = 2$, the encoder codes for decoder 1, resulting, after binning, in an index I_1 for the codeword U_1^n . The encoder sends out the modulo sum of the indices of the two codewords ($I_1 \oplus I_2$) along with the index of the action codeword. Decoder 1 has the X_i sequence when $A = 2$ and hence, it has the index I_2 . Therefore, it can recover its desired index I_1 from $I_1 \oplus I_2$. A similar analysis holds for decoder 2.

Example 2: Binary source with Hamming distortion and no cost constraint. Let $Y = \emptyset$ and $X \sim \text{Bern}(1/2)$. Assume no cost on the actions taken: $\Lambda(A = 1) = \Lambda(A = 2) = 0$ and let the distortion measure be Hamming. Then, the rate distortion trade-off evaluates to

$$\begin{aligned} R &= \min_{\alpha} \max \{ \alpha (1 - H_2(D_1/\alpha)) \mathbf{1}(D_1/\alpha \leq 1/2), \\ &\quad (1 - \alpha) (1 - H_2(D_2/(1 - \alpha))) \mathbf{1}(D_2/(1 - \alpha) \leq 1/2) \}, \end{aligned}$$

where $\mathbf{1}(x)$ denotes the indicator function. As a check, note that if $D_1, D_2 \rightarrow 0$, then the rate obtained is $1/2$, which agrees with the rate obtained in Corollary 1 for the lossless case. The result follows from explicitly evaluating the result in Proposition 2. Let $P(A = 2) = \alpha$. From Proposition 2, we have

$$\begin{aligned} R &\geq I(X; A) + P(A = 2)I(X; U_1|Y, A = 2) \\ &= 1 - (1 - \alpha)H(X|A = 1) - \alpha H(X|A = 2) + \alpha H(X|A = 2) - \alpha H(X|U_1, A = 2) \\ &\geq \alpha - \alpha H(X|U_1, A = 2) \\ &\geq \alpha(1 - H(X \oplus \hat{X}_1|U_1, A = 2)) \\ &\geq \alpha \left(1 - H_2\left(\frac{D_1}{\alpha}\right) \right) \mathbf{1}\left(\frac{D_1}{\alpha} \leq 1/2\right). \end{aligned}$$

The last step follows from the observations that (i) if $D_1/\alpha > 1/2$, then we lower bound R by 0; and (ii) if $D_1/\alpha \leq 1/2$, then from the distortion constraint $\alpha \mathbb{E} d(X, \hat{X}_1|A = 2) \leq D_1$, $H(X \oplus \hat{X}_1|A = 2) \leq H_2(D_1/\alpha)$.

The other bound is derived in the same manner. The fact that this rate can be attained is straightforward, since we can choose $U_1 = \hat{X}_1$ when $A = 2$ and $U_2 = \hat{X}_2$ when $A = 1$. In this example, the action sequence is independent of the source, but unlike the case of lossless source coding, $P(A = 1)$ is not in general equal to $P(A = 2)$. It depends on the distortion constraints for the individual decoders. A surface plot of the rate versus distortion constraints for the two decoders is shown in Figure 6.

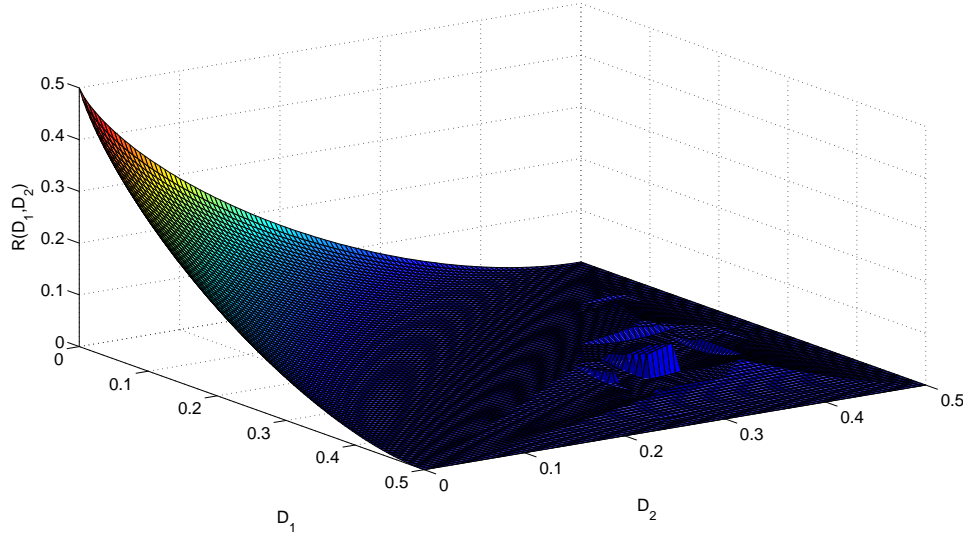


Fig. 6: Plot of rate versus distortions. The figure above plots the rate distortion surface $R(D_1, D_2)$ for the Example 2. There is no side information, i.e., $Y = \emptyset$ and $X \sim \text{Bern}(1/2)$. Assume no cost on the actions taken: $\Lambda(A = 1) = \Lambda(A = 2) = 0$ and let the distortion measure be Hamming. Note that if any of $D_1, D_2 \rightarrow 0.5$, R approaches 0, also if $D_1 = D_2 = 0$, rate is 0.5

C. Connections with Complementary Delivery

In the prequel, we consider several cases for switching dependent side information in which the achievability scheme has a simple “modulo sum” interpretation for the terms in the max function. This interpretation is not unique to our setup and in this subsection, we consider the complementary delivery setting [7] in which this interpretation also arises. Formally, the complementary delivery problem is a special case of our setting and is obtained by letting $A = \emptyset$, $X = (\tilde{X}, \tilde{Y})$, $P(Y_1, Y_2|X) = 1_{Y_1=\tilde{X}, Y_2=\tilde{Y}}$, $\Lambda(A) = 0$, $d_1(X, \hat{X}_1) = d'_1(\tilde{Y}, \hat{X}_1)$ and $d_2(X, \hat{X}_2) = d'_2(\tilde{X}, \hat{X}_2)$. For this subsection, for notational convenience, we will use X in place of \tilde{X} , Y in place of \tilde{Y} , \hat{Y} in place of \hat{X}_1 and \hat{X} in place of \hat{X}_2 . This setting is shown in Figure 7. In [7], the following achievable rate was established

$$R(D_1, D_2) \geq \max\{I(U; Y|X), I(U; X|Y)\}, \quad (3)$$

for some $p(u|x, y)$ satisfying $E d_1(Y, \hat{Y}(U, X)) \leq D_1$ and $E d_2(X, \hat{X}(U, Y)) \leq D_2$.

Our achievability scheme in Theorem 3 generalizes this scheme when specialized to the complementary delivery setting, but we do not yet know if our achievable rate can be strictly smaller for the same distortions. However, by taking a modulo sum interpretation for the terms in the $\max\{\cdot\}$ function in (3), as we have done for several examples in this paper, we are able to give simple proofs and explicit characterization for two canonical cases: the Quadratic Gaussian and the doubly symmetric binary Hamming distortion complementary delivery problems. While characterizations for these two settings also appear independently in [12], our approach in characterizing these settings is different from that in [12], and we believe would be of interest to readers. Furthermore, by taking the “modulo sum” interpretation, we establish the following, which may be a useful observation in practice: “For

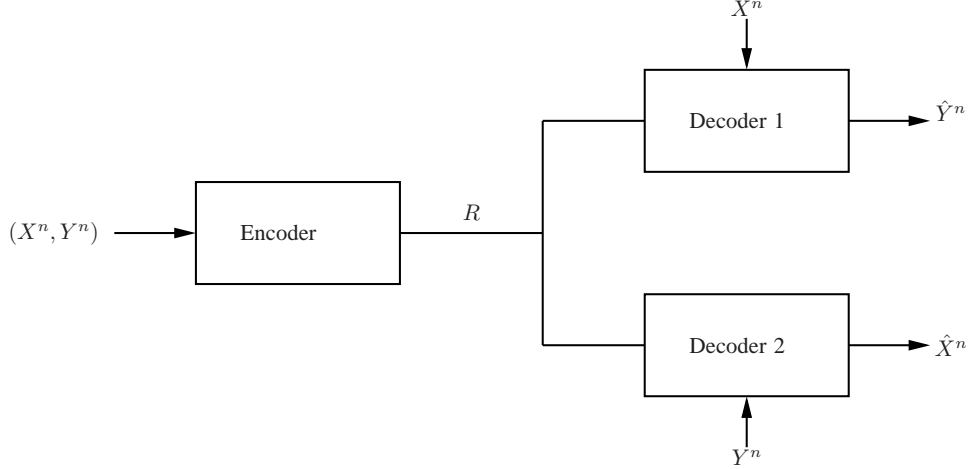


Fig. 7: Complementary Delivery setting

the Quadratic Gaussian complementary delivery problem, if one has a good code (in the sense of achieving the optimum rate distortion tradeoff) for the point to point Wyner-Ziv [1] Quadratic Gaussian setup, then a simple modification exists to turn the code into a good code for the Quadratic Gaussian complementary delivery problem.” A similar observation holds for the doubly symmetric binary Hamming distortion case. We first consider the Quadratic Gaussian case.

Proposition 3: Quadratic Gaussian complementary delivery. Let $Y = X + Z$, where $Z \sim N(0, N)$ is independent of $X \sim N(0, P)$, and the distortion measures be mean square distortion. Let $P' = PN/(P+N)$. The rate distortion region for the non-trivial constraints of $D_2 \leq P'$ and $D_1 \leq N$ is given by

$$R(D_1, D_2) = \max \left\{ \frac{1}{2} \log \left(\frac{N}{D_1} \right), \frac{1}{2} \log \left(\frac{P'}{D_2} \right) \right\}.$$

Proof:

Converse

The converse follows from straightforward cutset bound arguments. The reader may notice that the expression given above is the maximum of the Quadratic Gaussian Wyner-Ziv [1] rate to decoder 1 and the Quadratic Gaussian Wyner-Ziv rate to decoder 2, or equivalently the maximum of the two cutset bounds. Clearly, this rate is the lowest possible for the given distortions.

Achievability

We now show that it is also achievable using a modulo sum interpretation for (3). Consider first encoding for decoder 1. From the Quadratic Gaussian Wyner-Ziv result, we know that side information at the encoder is redundant. Therefore, without loss of optimality, the encoder can code for decoder 1 using only Y^n , resulting in the codeword U_Y^n and the corresponding index I_Y after binning. Similarly, for decoder 2, the encoder can code for decoder 2 using X^n only, resulting in the codeword U_X^n and index I_X after binning. The encoder then sends out the index $I_X \oplus I_Y$. Since decoder 1 has the X^n sequence as side information, it knows the index I_X and can therefore recover I_Y from $I_X \oplus I_Y$. The same decoding scheme works as well for decoder 2. Therefore, we have shown the achievability of the given rate expression. We note further that this scheme corresponds to setting $U = (U_X, U_Y)$ such that $U_X - X - Y - U_Y$ in rate expression (3). ■

Remark 4.4: As shown in our proof of achievability, if we have a good practical code for the Wyner-Ziv Quadratic Gaussian problem, then we also have a good practical code for the complementary delivery problem setting. We first develop two point to point codes: one for the Wyner-Ziv Quadratic Gaussian case with X as the source and Y as the side information, and another for the case where Y is the source and X is the side information. A good code for the complementary delivery setting is then obtained by taking the modulo sum of the indices produced by these two point to point codes.

We now turn to the doubly symmetric binary sources with Hamming distortion case. Here, the achievability scheme involves taking the modulo sum of the sources X^n and Y^n .

Proposition 4: Doubly symmetric binary source with Hamming distortion. Let $X \sim \text{Bern}(1/2)$, $Y \sim \text{Bern}(1/2)$, $X \oplus Y \sim \text{Bern}(p)$ and both distortion measures be Hamming distortion. Assume, without loss of generality, that $D_1, D_2 \leq p$. Then,

$$R(D_1, D_2) = \max\{H(p) - H(D_1), H(p) - H(D_2)\}.$$

Proof: The converse again follows from straightforward cutset bounds by considering decoders 1 and 2 individually. For the achievability scheme, let $Z = X \oplus Y$ and assume that $D_1 \leq D_2$. Since Z is i.i.d. $\text{Bern}(p)$, using a point to point code for Z at distortion D_1 , we obtain a rate of $H(p) - H(D_1)$. Denote the reconstruction for Z at time i by \hat{Z}_i . Decoder 1 reconstructs Y_i by $\hat{Y}_i = X_i \oplus \hat{Z}_i$ for $i \in [1 : n]$. Similarly, decoder 2 reconstructs X by $\hat{X}_i = Y_i \oplus \hat{Z}_i$ for $i \in [1 : n]$. To verify that the distortion constraint holds, note that $d_1(Y_i, X_i \oplus \hat{Z}_i) = Y_i \oplus X_i \oplus \hat{Z}_i = Z_i \oplus \hat{Z}_i$. Since \hat{Z} is a code that achieves distortion D_1 , \hat{Y} satisfies the distortion constraint for decoder 1. The same analysis holds for decoder 2. ■

Remark 4.5: In this case, we only need a good code for the standard point to point rate distortion problem for a binary source. A good rate distortion code for a binary source is also a good code for the doubly symmetric binary source with Hamming distortion complementary delivery problem.

Remark 4.6: In our scheme, the reconstruction symbols at time i depend only on the received message and the side information at the decoder at time i . Therefore, for this case, the rate distortion region for causal reconstruction [8] is the same as the rate distortion region for noncausal reconstruction.

V. ACTIONS TAKEN AT THE ENCODER

We now turn to the case where the encoder takes action (figure 3) instead of the decoders. When the actions are taken at the encoder, the general rate-cost-distortion tradeoff region is open even for the case of a single decoder. Special cases which have been characterized includes the lossless case [2]. In this section, we consider a special case of lossless source coding with K decoders in which we can characterize the rate-cost tradeoff region.

Theorem 4: Special case of lossless source coding with actions taken at the encoder. Let the action channel be given by the conditional distribution $P_{Y_1, Y_2, \dots, Y_K | X, A}$. Assume further that $A = f_1(Y_1) = f_2(Y_2) = \dots, f_K(Y_K)$. Then, the minimum rate required for lossless source coding with actions taken at the encoder and cost constraint C is given by

$$R = \min \left[\max_{j \in [1:K]} \{H(X|Y_j, A)\} - H(A|X) \right]^+,$$

where minimization is over the joint distribution $p(x)p(a|x)p(y_1, y_2, \dots, y_K|x, a)$ such that $\mathbb{E} \Lambda(A) \leq C$.

Proof:

Converse The proof of converse is a straightforward extension from the single decoder case given in [2]. We give the proof here for completeness. Consider the rate required for decoder j .

$$\begin{aligned} nR &\geq H(M) \\ &\geq H(M, X^n | Y_j^n) - H(X^n | M, Y_j^n) \\ &\geq H(M, X^n | Y_j^n) - n\epsilon_n \\ &\stackrel{(a)}{=} H(X^n | Y_j^n) - n\epsilon_n \\ &\stackrel{(b)}{=} H(X^n) + H(Y_j^n | X^n, A^n) - H(Y_j^n) - n\epsilon_n \\ &\geq \sum_{i=1}^n (H(X_i) + H(Y_{ji} | X_i, A_i) - H(Y_{ji})) - n\epsilon_n, \end{aligned}$$

where (a) follows from the fact that M is a function of X^n and (b) follows from A^n being a function of X^n . The last step follows from X^n being a discrete memoryless source; the action channel being memoryless and

conditioning reduces entropy. As before, we define Q to be an uniform random variable over $[1 : n]$ independent of all other random variables to obtain

$$\begin{aligned} R &\geq H(X) + H(Y_j|X, A) - H(Y_j) - \epsilon_n \\ &= H(X) + H(Y_j, X|A) - H(X|A) - H(Y_j) - \epsilon_n \\ &= H(X|A, Y_j) + I(X; A) - I(Y_j; A) - \epsilon_n \\ &= H(X|A, Y_j) - H(A|X) - \epsilon_n. \end{aligned}$$

The last step follows from the fact that $A = f_j(Y_j)$. Combining the lower bounds over K decoders then give us the achievable rate stated in the Theorem.

Achievability We give a sketch of achievability since the techniques used are relatively straightforward. Assume first that $R > 0$. We first bin the set of X^n sequences to $2^{n(\max_{j \in [1:K]} H(Y_j|X, A) + \epsilon)}$, $\mathcal{B}(M_X)$, $M_X \in [1 : 2^{n(\max_{j \in [1:K]} H(Y_j|X, A) + \epsilon)}]$. Given an x^n sequence, we first find the bin index m_x such that $x^n \in \mathcal{B}(m_x)$. We then split m_x into two sub-messages: $m_{xr} \in [1 : 2^{\max_{j \in [1:K]} \{H(X|Y_j, A)\} - H(A|X) + 2\epsilon}]$ and $m_{xa} \in [1 : 2^{n(H(A|X) - \epsilon)}]$. m_{xr} is transmitted over the noiseless link, giving us the rate stated in the Theorem. As for m_{xa} , we will send the message through the action channel by treating the action channel as a channel with i.i.d. state X noncausally known at the transmitter (A). We can therefore use Gel'fand Pinsker coding [13] for this channel.

Each decoder first decodes m_{xa} from their side information Y_j . From the condition that $A = f_j(Y_j)$ for all j , we have $H(A|X) - \epsilon = I(Y_j; A) - I(X; A) - \epsilon$. From analysis of Gel'fand-Pinsker coding, since $|\mathcal{M}_{xa}| = I(Y_j; A) - I(X; A) - \epsilon$, the probability of error in decoding m_{xa} goes to zero as $n \rightarrow \infty$. The decoder then reconstructs m_x from m_{xr} and m_{xa} . It then finds the unique $\hat{x}^n \in \mathcal{B}(m_x)$ that is jointly typical with Y_j^n and A^n . Note that due to Gel'fand-Pinsker coding, the true x^n sequence is jointly typical with Y_j^n and A^n with high probability. Therefore, the probability of error in this decoding step goes to zero as $n \rightarrow \infty$ since we have $2^{n(\max_{j \in [1:K]} H(Y_j|X, A) + \epsilon)}$ bins.

For the case where $R = 0$, we send the entire message through the action channel. ■

Example 3: Consider the case of $K = 2$ with switching dependent side information: $A = \{1, 2\}$ and $(X, Y) \sim p(x, y)$ with $P_{Y_1, Y_2|X, A}$ specified by $Y_1 = Y, Y_2 = e$ when $A = 1$ and $Y_1 = e, Y_2 = Y$ when $A = 2$. Note that A is a function of Y_1 , and also of Y_2 . It therefore satisfies the condition in Theorem 4. Let $P(A = 1) = \alpha$, $\Lambda(A = 1) = C_1$ and $\Lambda(A = 2) = C_2$. The rate-cost tradeoff is characterized by

$$\begin{aligned} R = \max\{ &\alpha H(X|A = 1, Y) + (1 - \alpha)H(X|A = 1), (1 - \alpha)H(X|A = 2, Y) + \alpha H(X|A = 2) \\ &+ H(X) - H_2(\alpha) - \alpha H(X|A = 1) - (1 - \alpha)H(X|A = 2) \} \end{aligned}$$

for some $p(a|x)$ satisfying $\alpha C_1 + (1 - \alpha)C_2 \leq C$.

VI. OTHER SETTINGS

In this section, we consider other settings involving multi-terminal source coding with action dependent side information. The first setting that we consider in this section generalizes [2, Theorem 7] to the case where there is a rate-limited link from the source encoder to the action encoder. The second setting we consider is a case of successive refinement with actions.

A. Single decoder with Markov Form $X-A-Y$ and rate limited link to action encoder

In this subsection, we consider the setting illustrated in Figure 8. Here, we have a single decoder with actions taken at an action encoder. The source encoder have access to source X^n and sends out two indices $M \in [1 : 2^{nR}]$ and $M_A \in [1 : 2^{nR_A}]$. The action encoder is a function $f : M_A \rightarrow A^n$. In addition, we have the Markov relation $X - A - Y$. That is, the side information Y is dictated only by the action A taken. The other definitions remain the same and we omit them here.

Proposition 5: $R(D, C)$ for the setting shown in figure 8 is given by

$$R(D, C) = \min \max\{I(X; \hat{X}) - R_A, I(X; \hat{X}) - I(A; Y)\},$$

where the minimization is over $p(x)p(a)p(y|a)p(\hat{x}|x)$ satisfying the cost and distortion constraints $E d(X, \hat{X}) \leq D$ and $E \Lambda(A) \leq C$.

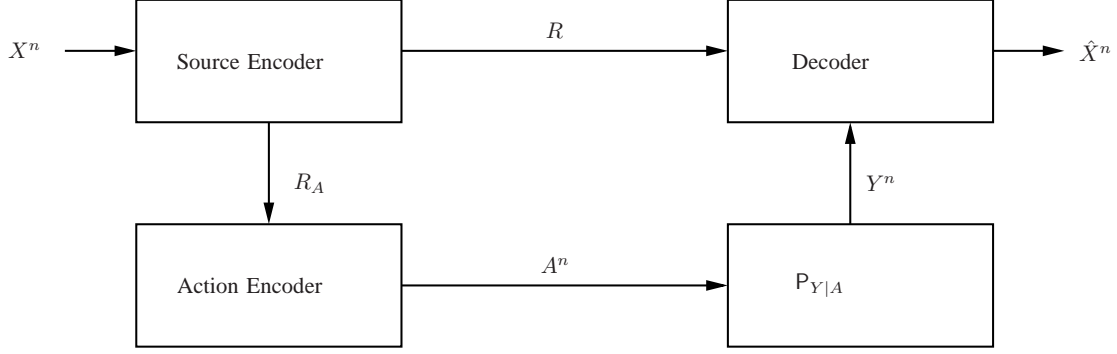


Fig. 8: Lossy source coding with rate limited link to action encoder

Remark 6.1: If we set $R_A = \infty$ in Proposition 5, then we recover the result in [2, Theorem 7]. Essentially, the source encoder tries to send as much information as possible through the rate limited action link until the link saturates.

Proof:

Achievability: The achievability is straightforward. Using standard rate distortion coding, we cover X^n with $2^{n(I(X;\hat{X})+\epsilon)}$ \hat{X}^n codewords. Given a source sequence x^n , we find an \hat{X}^n that is jointly typical with x^n . We then split the index M_X corresponding to the chosen \hat{X}^n codeword into two parts: $M_A \in [1 : 2^{n(\min\{R_A, I(A;Y)\}+\epsilon)}]$ and $M \in [1 : 2^{nR}]$. The action encoder takes the index and transmit it through the action channel. Since the rate of M_A is less than $I(A;Y) - \epsilon$, the decoder can decode M_A with high probability of success. It then combines M_A with M to obtain the index of the reconstruction codeword \hat{X}^n .

Converse Given a code that satisfy the distortion and cost constraints, we have

$$\begin{aligned}
 nR &\geq H(M) \\
 &= I(X^n; M) \\
 &\geq I(X^n; M) - I(X^n; Y^n) \\
 &= I(X^n; \hat{X}^n) - I(X^n, M_A; Y^n) \\
 &\stackrel{(a)}{\geq} \sum_{i=1}^n I(X_i; \hat{X}_i) - I(X^n, M_A, A^n; Y^n) \\
 &\stackrel{(b)}{\geq} \sum_{i=1}^n I(X_i; \hat{X}_i) - I(M_A, A^n; Y^n).
 \end{aligned}$$

(a) follows from the fact that A^n is a function of M_A . (b) follows from the Markov chain $X - A - Y$. Now, it is easy to see that $I(M_A, A^n; Y_i) \leq \min\{nR_A, \sum_{i=1}^n I(A_i; Y_i)\}$. The bound on the rate is then single letterized in the usual manner, giving us

$$R(D, C) = \min \max \{I(X; \hat{X}) - R_A, I(X; \hat{X}) - I(A; Y)\},$$

for some $p(a, \hat{x}|x)$ satisfying the distortion and cost constraints. Finally, we note that $p(a, \hat{x}|x)$ can be restricted to the form $p(a)p(\hat{x}|x)$. To see this, note that none of the terms depend on the joint $p(a, \hat{x}|x)$. Furthermore, due to the Markov condition $X - A - Y$, it suffices to consider A independent of X , giving us the p.m.f in the Proposition. ■

B. Successive refinement with actions

The next setup that we consider is a case of successive refinement [14], [15] with actions taken at the “more capable” decoder. The setting is shown in Figure 9.

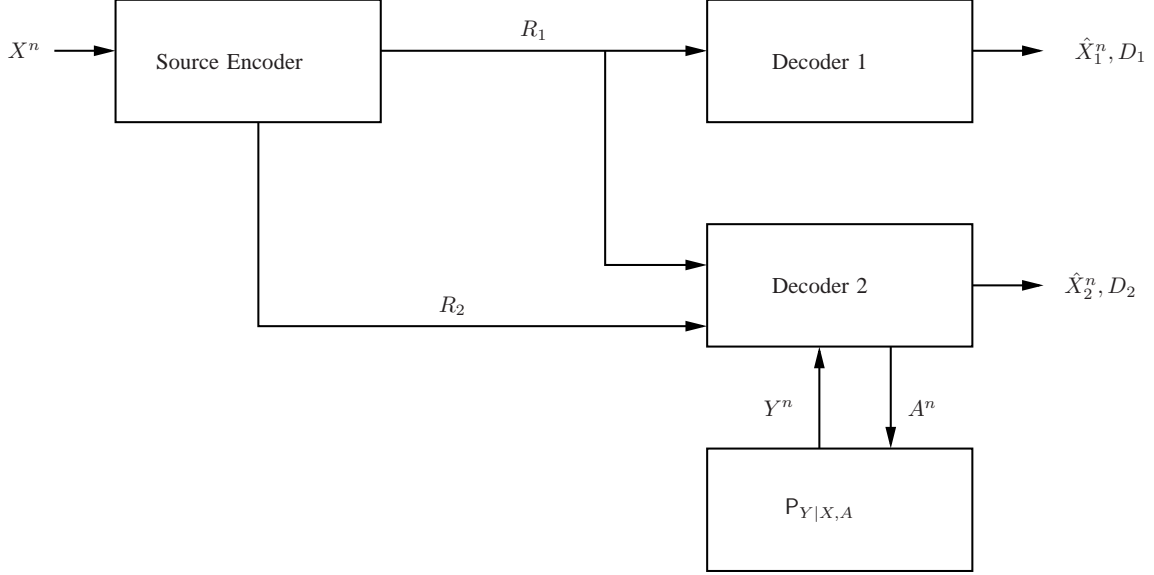


Fig. 9: Successive refinement with actions

Proposition 6: Successive refinement with actions taken at the more capable decoder For the setting shown in figure 9, the rate distortion cost tradeoff region is given by

$$\begin{aligned} R_1 &\geq I(X; \hat{X}_1), \\ R_1 + R_2 &\geq I(X; \hat{X}_1, A) + I(X; U | \hat{X}_1, Y, A) \end{aligned}$$

for some $p(\hat{x}_1, a, u|x)$ satisfying

$$\begin{aligned} \mathbb{E} d_1(X, \hat{X}_1) &\leq D_1, \\ \mathbb{E} d_1(X, \hat{X}_2(U, Y, A)) &\leq D_2, \\ \mathbb{E} \Lambda(A) &\leq C. \end{aligned}$$

The cardinality of the auxiliary U may be upper bounded by $|\mathcal{U}| \leq |\mathcal{X}||\hat{\mathcal{X}}_1||\mathcal{A}| + 1$.

If we restrict $R_2 = 0$, then Proposition 6 gives the rate-distortion-cost tradeoff region for a special case of Proposition 1. That is, the case when $Y_2 = \emptyset$ and actions are taken only at decoder 1.

Proof:

Achievability: We give the case where $R_1 = I(X; \hat{X}_1) + \epsilon$ and $R_2 = I(X; A | \hat{X}_1) + I(X; U | \hat{X}_1, Y, A) + 3\epsilon$. The general region stated in the Proposition can then be obtained by rate splitting of R_2 .

Codebook generation

- Generate 2^{nR_1} $\hat{X}_1^n(m_1)$ sequences according to $\prod_{i=1}^n p(\hat{x}_{1i})$, $m_1 \in [1 : 2^{nR_1}]$.
- For each $\hat{X}_1^n(m_1)$ sequence, generate $2^{n(I(X; A | \hat{X}_1) + \epsilon)}$ $A^n(m_1, m_{21})$ sequences according to $\prod_{i=1}^n p(a_i | \hat{x}_{1i})$.
- For each $\hat{X}_1^n(m_1)$ and $A^n(m_1, m_2)$ sequence pair, generate $2^{n(I(X; U | \hat{X}_1, A) + \epsilon)}$ $U^n(m_1, m_{21}, l_{22})$ sequences according to $\prod_{i=1}^n p(u_i | \hat{x}_{1i}, a_i)$.
- Partition the set of l_{22} indices into $2^{I(X; U | \hat{X}_1, Y, A) + 2\epsilon}$ bins, $\mathcal{B}(m_1, m_{21}, m_{22})$, $m_{22} \in [1 : 2^{n(I(X; U | \hat{X}_1, Y, A) + 2\epsilon)}]$.

Encoding

- Given a sequence x^n , the encoder first looks for an $\hat{x}_1^n(m_1)$ sequence such that $(x^n, \hat{x}_1^n) \in \mathcal{T}_\epsilon^{(n)}$. This step succeeds with high probability since $R_1 = I(X; \hat{X}_1) + \epsilon$.

- Next, the encoder looks for an $A^n(m_1, m_{21})$ sequence such that $(x^n, a^n, \hat{x}_1^n) \in aep$. This step succeeds with high probability since we have $2^{n(I(X;A|\hat{X}_1)+\epsilon)}$ A^n sequences.
- The encoder then looks for an $U^n(m_1, m_{21}, l_{22})$ sequence such that $(x^n, a^n, \hat{x}_1^n, u^n) \in aep$. This step succeeds with high probability since we have $2^{n(I(X;U|\hat{X}_1,A)+\epsilon)}$ U^n sequences.
- It then finds the bin index such that $l_{22} \in \mathcal{B}(m_1, m_{21}, m_{22})$.
- The encoder sends out the indices m_1 over the link R_1 and m_{21} and m_{22} over the link R_2 , giving us the stated rates.

Decoding and reconstruction

- Since decoder 1 has index m_1 , it reconstructs x^n using $\hat{x}_1(m_1)^n$. Since (x^n, \hat{x}_1^n) are jointly typical with high probability, the expected distortion satisfies the D_1 distortion constraint to within ϵ .
- For decoder 2, from m_1 and m_{21} , it recovers the action sequence $a^n(m_1, m_{21})$. It then takes the action $a^n(m_1, m_{21})$ to obtain its side information Y^n . With the side information, it recovers the u^n sequence by looking for the unique $\hat{l}_{22} \in \mathcal{B}(m_1, m_{21}, m_{22})$ such that $(u^n(m_1, m_{21}, \hat{l}_{22}), \hat{x}_1^n, a^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}$. Since there are only $2^{n(I(U;Y|\hat{X}_1,A)-\epsilon)}$ U^n sequences in the bin and $(u^n(m_1, m_{21}, l_{22}), \hat{x}_1^n, a^n, Y^n) \in \mathcal{T}_\epsilon^{(n)}$ with high probability from the fact that Y is generated i.i.d. according to $p(y|a_i, x_i)$, the probability of error goes to zero as $n \rightarrow \infty$. Decoder 2 then reconstructs x^n using $\hat{x}_{2i}(a_i, u_i, y_i)$ for $i \in [1 : n]$.

Converse: We consider only the lower bound for $R_1 + R_2$. The lower bound for R_1 is straightforward. Given a code which satisfies the distortion and cost constraints, we have

$$\begin{aligned}
n(R_1 + R_2) &\geq H(M_1, M_2) \\
&= H(M_1, M_2, A^n, \hat{X}_1^n) \\
&= H(A^n, \hat{X}_1^n) + H(M_1, M_2 | A^n, \hat{X}_1^n) \\
&\geq I(X^n; A^n, \hat{X}_1^n) + H(M_1, M_2 | A^n, \hat{X}_1^n, Y^n) - H(M_1, M_2 | Y^n, A^n, \hat{X}_1^n, X^n) \\
&= I(X^n; A^n, \hat{X}_1^n) + I(X^n; M_1, M_2 | A^n, \hat{X}_1^n, Y^n) \\
&= I(X^n; A^n, \hat{X}_1^n) + H(X^n | A^n, \hat{X}_1^n, Y^n) - H(X^n | A^n, \hat{X}_1^n, Y^n, M_1, M_2) \\
&= I(X^n; A^n, \hat{X}_1^n) + H(X^n, Y^n | A^n, \hat{X}_1^n) - H(Y^n | \hat{X}_1^n, A^n) - H(X^n | A^n, \hat{X}_1^n, Y^n, M_1, M_2) \\
&= H(X^n) - H(X^n | A^n, \hat{X}_1^n) + H(X^n, Y^n | A^n, \hat{X}_1^n) - H(Y^n | \hat{X}_1^n, A^n) - H(X^n | A^n, \hat{X}_1^n, Y^n, M_1, M_2) \\
&= H(X^n) + H(Y^n | X^n, A^n, \hat{X}_1^n) - H(Y^n | \hat{X}_1^n, A^n) - H(X^n | A^n, \hat{X}_1^n, Y^n, M_1, M_2) \\
&\geq \sum_{i=1}^n (H(X_i) + H(Y_i | X^n, A^n, \hat{X}_1^n, Y^{i-1}) - H(Y_i | \hat{X}_1^n, A^n, Y^{i-1}) - H(X_i | A^n, \hat{X}_1^n, Y^n, M_1, M_2)) \\
&\stackrel{(a)}{\geq} \sum_{i=1}^n (H(X_i) + H(Y_i | X_i, A_i, \hat{X}_{1i}) - H(Y_i | \hat{X}_{1i}, A_i) - H(X_i | A^n, \hat{X}_1^n, Y^n, M_1, M_2)) \\
&= \sum_{i=1}^n (H(X_i) + H(Y_i | X_i, A_i, \hat{X}_{1i}) - H(Y_i | \hat{X}_{1i}, A_i) - H(X_i | A^n, \hat{X}_1^n, Y^n, M_1, M_2)) \\
&\geq \sum_{i=1}^n (H(X_i) + H(Y_i | X_i, A_i, \hat{X}_{1i}) - H(Y_i | \hat{X}_{1i}, A_i) - H(X_i | U_i, A_i, Y_i, \hat{X}_{1i}))
\end{aligned}$$

(a) follows from the Markov Chain $(X^{n \setminus i}, A^{n \setminus i}, \hat{X}_1^{n \setminus i}, Y^{i-1}) - (\hat{X}_i, X_i, A_i) - Y_i$ and the last step follows from defining $U_i = (M_1, M_2, Y^{n \setminus i}, A^{n \setminus i})$. The proof is then completed in the usual manner by defining the time sharing uniform random variable Q and $U = (U_Q, Q)$, giving us

$$\begin{aligned}
R_1 + R_2 &\geq H(X) + H(Y | X, A, \hat{X}_1) - H(Y | \hat{X}_1, A) - H(X | U, A, Y, \hat{X}_1) \\
&= I(X; \hat{X}_1, A) + I(X; U | \hat{X}_1, Y, A).
\end{aligned}$$

The fact that \hat{X}_2 is a function of U , Y and A , which is straightforward. Finally, the cardinality bound on U may be obtained from standard techniques. Note that we need $|\hat{\mathcal{X}}_1||\mathcal{X}||\mathcal{A}| - 1$ letters to preserve $p(u, a, x)$ and two more to preserve the rate and distortion constraints. ■

Remark 6.2: An interesting question to explore characterizing the more general case when degraded side information is also available at decoder 1. That is, we have the side informations Y_1 at decoder 1 and Y_2 at decoder 2 are generated by a discrete memoryless channel $P_{Y_1, Y_2|X, A}$ such that $(X, A) - (Y_2, A) - (Y_1, A)$. This generalized setup would allow us to generalize Proposition 1 entirely and also leads to a generalization of successive refinement for the Wyner-Ziv problem in [16] to the action setting.

VII. CONCLUSION

In this paper, we considered an important class of multi-terminal source coding problems, where the encoder sends the description of the source to the decoders, which then take cost-constrained actions that affect the quality or availability of side information. We computed the optimum rate region for lossless compression, while for the lossy case we provide a general achievability scheme that is shown to be optimal for a number of special cases, one of them being the generalization of *Heegard-Berger-Kaspi* setting. (cf. [10], [11]). In all these cases in addition to a standard achievability argument, we also provided a simple scheme which has a *modulo sum* interpretation. The problem where the encoder takes actions rather than the decoders, was also considered. Finally, we extended the scope to additional multi-terminal source coding problems such as successive refinement with actions.

ACKNOWLEDGMENT

We thank Professor Haim Permuter and Professor Yossef Steinberg for helpful discussions. The authors' research was partially supported by NSF grant CCF-0729195 and the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-0939370; the Scott A. and Geraldine D. Macomber Stanford Graduate Fellowship and the Office of Technology Licensing Stanford Graduate Fellowship.

REFERENCES

- [1] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, 1976.
- [2] H. Permuter and T. Weissman, "Source coding with a side information "vending machine"," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4530–4544, July 2011.
- [3] H. Asnani, H. H. Permuter, and T. Weissman, "Probing capacity," *CoRR*, vol. abs/1010.1309, 2010.
- [4] K. Kittichokechai, T. J. Oechtering, M. Skoglund, and R. Thobaben, "Source and channel coding with action-dependent partially known two-sided state information," in *Proc. IEEE International Symposium on Information Theory*, 2010, pp. 629–633–194.
- [5] A. El Gamal and Y. H. Kim, "Lectures on network information theory," 2010, available online at ArXiv.
- [6] T. M. cover, "A proof of the data compression theorem of slepian and wolf for ergodic sources," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 226–228, 1975.
- [7] A. Kimura and T. Uyematsu, "Multiterminal source coding with complementary delivery," in *Proc. International Symposium on Information Theory and its Applications (ISITA)*, 2006, pp. 189–194.
- [8] T. Weissman and A. El Gamal, "Source coding with limited-look-ahead side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5218–5239, 2006.
- [9] H. G. Eggleston, *Convexity*. Cambridge: Cambridge University Press, 1958.
- [10] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 727–734, 1985.
- [11] A. H. Kaspi, "Rate-distortion function when side-information may be present at the decoder,," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 2031–2034, 1994.
- [12] R. Timo, A. Grant, and G. Kramer, "Rate-distortion functions for source coding with complementary side information," in *Proc. IEEE International Symposium on Information Theory*, St. Petersburg, Russia, 2011, pp. 2934–2938.
- [13] S. I. Gel'fand and M. S. Pinsker, "Coding for channels with random parameters," *Probl. Control and Information Theory*, vol. 9, pp. 19–31, 1980.
- [14] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 269–275, 1991.
- [15] —, "Successive refinement of information: Characterization of the achievable rates,," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 253–259, 1994.
- [16] Y. Steinberg and N. Merhav, "On successive refinement for the wyner-ziv problem," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1636–1654, 2004.

APPENDIX A

ACHIEVABILITY SKETCH FOR THEOREM 3

Codebook generation

- Generate $2^{n(I(X;A)+\epsilon)}$ $A^n(l_a)$, $l_a \in [1 : 2^{n(I(X;A)+\epsilon)}]$, sequences according to $\prod_{i=1}^n p(a_i)$.
- For each A^n sequence, generate $2^{n(I(U;X|A)+\epsilon)}$ $U^n(l_a, l_0)$, $l_0 \in [1 : 2^{n(I(U;X|A)+\epsilon)}]$, sequences according to $\prod_{i=1}^n p_{U|A}(u_i|a_i)$.
- Partition the set of indices corresponding to the U^n codewords uniformly to $2^{n(\max\{I(X;U|A,Y_1), I(X;U|A,Y_2)\}+2\epsilon)}$ bins, $\mathcal{B}_U(l_a, m_0)$, $m_0 \in [1 : 2^{n(\max\{I(X;U|A,Y_1), I(X;U|A,Y_2)\}+2\epsilon)}]$.
- For each pair of A^n and U^n sequences, generate $2^{n(I(V_1;X|A,U)+\epsilon)}$ $V_1^n(l_a, l_0, l_1)$, $l_1 \in [1 : 2^{n(I(V_1;X|A,U)+\epsilon)}]$, sequences according to $\prod_{i=1}^n p_{V_1|A,U}(v_{1i}|a_i, u_i)$.
- Partition the set of indices corresponding to the V_1^n codewords uniformly to $2^{n(I(X;V_1|U,A,Y_1)+2\epsilon)}$ bins, $\mathcal{B}_{V_1}(l_a, l_0, m_1)$, $m_1 \in [1 : 2^{n(I(X;V_1|U,A,Y_1)+2\epsilon)}]$.
- For each pair of A^n and U^n sequences, generate $2^{n(I(V_2;X|A,U)+\epsilon)}$ $V_2^n(l_a, l_0, l_2)$, $l_2 \in [1 : 2^{n(I(V_2;X|A,U)+\epsilon)}]$, sequences according to $\prod_{i=1}^n p_{V_2|A,U}(v_{2i}|a_i, u_i)$.
- Partition the set of indices corresponding to the V_2^n codewords uniformly to $2^{n(I(X;V_2|U,A,Y_2)+2\epsilon)}$ bins, $\mathcal{B}_{V_2}(l_a, l_0, m_2)$, $m_2 \in [1 : 2^{n(I(X;V_2|U,A,Y_2)+2\epsilon)}]$.

Encoding

- Given an x^n sequence, the encoder first looks for an $a^n(l_a)$ sequence such that $(x^n, a^n) \in \mathcal{T}_\epsilon^{(n)}$. If there is none, it outputs an index chosen uniformly at random from the set of possible l_a indices. If there is more than one, it outputs an index chosen uniformly at random from the set of feasible indices. Since there are $2^{n(I(X;A)+\epsilon)}$ such sequences, the probability of error $\rightarrow 0$ as $n \rightarrow \infty$.
- The encoder then looks for a $u^n(l_a, l_0)$ sequence that is jointly typical with $(a^n(l_a), x^n)$. If there is none, it outputs an index chosen uniformly at random from the set of possible l_0 indices. If there is more than one, it outputs an index chosen uniformly at random from the set of feasible indices. Since there are $2^{n(I(U;X|A)+\epsilon)}$ such sequences, the probability of error $\rightarrow 0$ as $n \rightarrow \infty$.
- Next, the encoder looks for a $v_1^n(l_a, l_0, l_1)$ sequence that is jointly typical with $(a^n(l_a), u^n(l_0), x^n)$. If there is none, it outputs an index chosen uniformly at random from the set of possible l_1 indices. If there is more than one, it outputs an index chosen uniformly at random from the set of feasible indices. Since there are $2^{n(I(V_1;X|A,U)+\epsilon)}$ such sequences, the probability of error $\rightarrow 0$ as $n \rightarrow \infty$.
- Next, the encoder looks for a $v_2^n(l_a, l_0, l_2)$ sequence that is jointly typical with $(a^n(l_a), u^n(l_0), x^n)$. If there is none, it outputs an index chosen uniformly at random from the set of possible l_2 indices. If there is more than one, it outputs an index chosen uniformly at random from the set of feasible indices. Since there are $2^{n(I(V_2;X|A,U)+\epsilon)}$ such sequences, the probability of error $\rightarrow 0$ as $n \rightarrow \infty$.
- The encoder then sends out the indices l_a, m_0, m_1 and m_2 such that $l_0 \in \mathcal{B}_U(l_a, m_0)$, $l_1 \in \mathcal{B}_{V_1}(l_a, l_0, m_1)$ and $l_2 \in \mathcal{B}_{V_2}(l_a, l_0, m_2)$.

Decoding and reconstruction

Decoder 1:

- Decoder 1 first takes the action sequence $a^n(l_a)$ to obtain the side information Y_1^n . We note that if $(a^n(l_a), x^n, u^n(l_a, l_0), v_1^n(l_a, l_0, l_1)) \in \mathcal{T}_\epsilon^{(n)}$, then $P\{(a^n(l_a), x^n, u^n(l_a, l_0), v_1^n(l_a, l_0, l_1), Y_1^n) \in \mathcal{T}_\epsilon^{(n)}\} \rightarrow 1$ as $n \rightarrow \infty$ by the conditional typicality lemma [5, Chapter 2] and the fact that $Y_1^n \sim \prod_{i=1}^n p(y_{1i}|x_i, a_i)$.
- Decoder 1 then decodes U^n . It does this by finding the unique \hat{l}_0 such that $u^n(l_a, \hat{l}_0) \in \mathcal{B}_U(l_a, m_0)$. If there is none or more than one such \hat{l}_0 , an error is declared. Following standard analysis for the Wyner-Ziv setup (see for e.g. [5, Chapter 12]), the probability of error goes to zero as $n \rightarrow \infty$ since there are less than or equal to $2^{n(I(U;Y_1|A)-\epsilon)}$ U^n sequences within each bin.
- Similarly, decoder 1 decodes V_1^n . It does this by finding the unique \hat{l}_1 such that $v_1^n(l_a, \hat{l}_0, \hat{l}_1) \in \mathcal{B}_{V_1}(l_a, \hat{l}_0, m_1)$. If there is none or more than one such \hat{l}_1 , an error is declared. As with the previous step, the probability of error goes to zero as $n \rightarrow \infty$ since there are only $2^{n(I(V_1;Y_1|A,U)-\epsilon)}$ V_1^n sequences within each bin.
- Decoder 1 then reconstructs x^n as $\hat{x}_{1i}(a_i(l_a), u_i(l_a, \hat{l}_0), v_{1i}(l_a, \hat{l}_0, \hat{l}_1), y_{1i})$ for $i \in [1 : n]$.

Decoder 2: As the decoding steps for decoder 2 are similar to that for 1, we will only mention the differences here. That is, decoder 2 uses side information Y_2^n instead of Y_1^n to perform the decoding operations and instead of decoding V_1^n , decoder 2 decodes V_2^n .

- Decoder 2 decodes V_2^n . It does this by finding the unique \hat{l}_2 such that $v_2^n(l_a, \hat{l}_0, \hat{l}_2) \in \mathcal{B}_{V_2}(l_a, \hat{l}_0, m_2)$. If there is none or more than one such \hat{l}_2 , an error is declared. As with the previous step, the probability of error goes to zero as $n \rightarrow \infty$ since there are only $2^{n(I(V_2; Y_2|A, U) - \epsilon)}$ V_2^n sequences within each bin.
- Decoder 1 then reconstructs x^n as $\hat{x}_{2i}(a_i(l_a), u_i(l_a, \hat{l}_0), v_{2i}(l_a, \hat{l}_0, \hat{l}_2), y_{2i})$ for $i \in [1 : n]$.

Distortion and cost constraints

- For the cost constraint, since the chosen A^n sequence is typical with high probability, $E \Lambda(A^n) \leq C + \epsilon$ by the typical average lemma [5, Chapter 2].
- For the distortion constraints, since the probability of “error” goes to zero as $n \rightarrow \infty$ and we are dealing only with finite cardinality random variables, following the analysis in [5, Chapter 3], we have

$$\begin{aligned} \frac{1}{n} E d_1(X^n, \hat{X}_1^n) &\leq D_1 + \epsilon, \\ \frac{1}{n} E d_2(X^n, \hat{X}_2^n) &\leq D_2 + \epsilon. \end{aligned}$$